Mean or median? Real life example.

TL;DR Use mean (average) to get an idea, and median to make an informed decision.

Why:

Suppose a survey asks me “Hey, Alina, how much cash do you usually carry each day in your wallet?” The question might come from a bank entering the local market, or from a business venue nearby.

At first I would think that each day of the week I have a different amount, so I wouldn’t be sure what to answer. Then I would realize there are two possible answers:

1. The average amount (by adding amounts of each day, and dividing the total by 7).
2. The amount that would have been average if it wouldn’t have been for atypical days (i.e. Sunday when I carry more than usual).

What would be the most useful answer for a business trying to make an informed decision about the services they want to provide?

If I write down the amount of cash in my wallet every day of the week (Monday first), we notice that the amount on Sunday is an outlier, a value that’s way off when compared to others:

Mo	Tu	We	Th	Fr	Sa	Su
10	5	5	20	15	20	60

To calculate the mean we add them all up, then divide the total by 7 and we get the result 19.3.

The interesting fact is that in five out of seven days, I have an amount lower than 19.3. Suppose the business is trying to sell me something priced at 16 euro thinking I sure have 19.3 in my pocket. Well, the sale has low chances of happening because I have enough money only on two days out of seven.

Let’s calculate the median and see what happens. We sort the values lowest to largest: 5, 10, 10, 15, 15, 20, 60. Then we pick the value in the middle and that’s our median: 15.

Notice that the median 15 is close to the planned price of 16 euro. Yes, I still have enough money only on two days out of seven. But if the business decides to trim a bit the price to 15 or 14.90 :), I will have enough money on four days out of seven. And this increases the chances for the business to sell the product.

By answering the survey using the median, the business gets a more valuable insight and has the opportunity to better design its sales strategy.

From my observations so far, median is a better indicator than mean when:

- Percentage of outliers from total sample is higher than expected.
- Distance between mean and outliers is greater than expected.
- Distribution of values is still gaussian, but skewed to one side.

Get the code for the graph in the header at https://github.com/alina-molnar/Mean_or_median/blob/main/mean_or_median.py

Anything to add? Let me know in the comments.